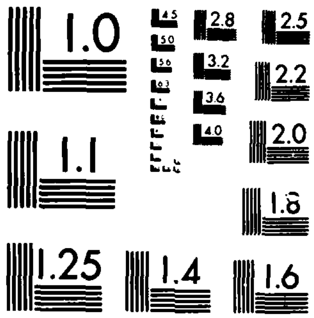
AD-A163 388 ON DENSITY ESTIMATION FROM CENSORED DATA BY PENALIZED 1/1
LIKELIHOOD METHODS..(U) JOHNS HOPKINS UNIV BALTIMORE MD
DEPT OF MATHEMATICAL SCIENCES.. V K KLONIAS ET AL.
UNCLASSIFIED JUN 85 TR-418 ONR-TR-85-3 N00014-79-C-0801 F/G 12/1 NL

END
FILMED
DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

AD-A163 388

(9)

# On Density Estimation from Censored Data by Penalized Likelihood Methods

Short Title: Density Estimation Under Censoring

by


V. K. Klonias

and

John C. Wierman







Mathematical Sciences Department
Johns Hopkins University
Baltimore, Maryland 21218



ONR Technical Report No. 85-3
Technical Report No. 418
June, 1985



AMS 1980 Subject Classification: Primary 62G05, Secondary 62A10, 62P10, 62N05.

Key Words and Phrases: maximum penalized likelihood estimation, density estimation, censored data, Kaplan-Meier estimator, hazard rate, spline function, reproducing kernel Hilbert space.

Research supported by the U.S. Department of Navy under Office of Naval Research Contract No. N00014-79-C-0801 and the National Science Foundation under Grant No. MCS-8303238. Reproduction in whole or in part is permitted for any purpose of the United States Government.

A

# On Density Estimation from Censored Data by Penalized Likelihood Methods


## Abstract

Estimators for the probability density function, cumulative distribution function, and hazard function are proposed in the random censorship setting. The estimators are derived from the Kaplan-Meier product limit estimator by maximum penalized likelihood methods. ~~We~~ The authors establish the existence and uniqueness of the estimates, which are exponential splines with knots at the uncensored observations, and provide an efficient algorithm for their numerical evaluation. ~~We~~ They prove the consistency, in probability and almost surely, of the density estimates in the Hellinger distance, the $L^p$ norms for $p = 1, 2, \infty$, and the Sobolev norm. The corresponding hazard rate estimator converges uniformly on bounded intervals.

## 1. Introduction

The classic problem in the independent random censorship model is to estimate the distribution function nonparametrically. The maximum likelihood estimator is the well-known product limit estimator introduced by Kaplan and Meier (1958). We propose an estimator of the density derived from the Kaplan-Meier estimator by maximum penalized likelihood techniques.

For uncensored data, the maximum penalized likelihood estimator (MPLE) was introduced by Good and Gaskins (1971, 1980). Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from a distribution F with density f, and let $F_n$ denote the corresponding empirical distribution function. The MPLE, denoted by $f_n$, maximizes the likelihood $\prod_{i=1}^{n} f(X_i)$ over a space of "smooth" functions. (Requiring smoothness avoids the Dirac delta solution of the unconstrained problem.) Equivalently, $f_n$ is the maximizer of

$$n \int \log f \, dF_n - \Phi(f) \tag{1.1}$$

subject to $\int f = 1$ and $f \geq 0$, where $\Phi(f)$ is a "roughness penalty". DeMontricher, Tapia, and Thompson (1975) rigorously established the existence and uniqueness of the solution $f_n$ within the framework of Sobolev spaces, and showed that the resulting MPLE is a spline function with knots at the sample points. Silverman (1982) proposed and studied the statistical properties and asymptotic distribution theory of a class of estimators with roughness penalties on log f. Klonias (1984) obtained existence, uniqueness and consistency results for a broad class of penalty functionals on $f^{\frac{1}{2}}$.

In the censored data setting, Lubecke and Padgett (1984) proposed estimating the density f by the maximizer of the penalized conditional likelihood, given which observations were censored. Questions of evaluation of the

estimator and consistency were not addressed.

We propose estimators for the density, distribution, and hazard functions derived by maximum penalized likelihood techniques. These estimators are based on an estimate of the root-density $v=f^{\frac{1}{2}}$ denoted by $u_n$, which is an exponential spline function with knots at the uncensored observations. The estimator $u_n$ corresponds to the "first MPLE" of Good and Gaskins (1971) in the uncensored setting. The advantages of parameterizing the problem through the root-density are that it is square-integrable, conveniently allowing the use of Hilbert space methods, and avoids the nonnegativity constraint $f \geq 0$, while providing the same density estimator as the direct approach - for the same penalty functional - when the MPLE $u_n$ turns out to be nonnegative, as is the case here; see Lemma 3.1 of DeMontricher et al (1975). In addition, the square root transformation is a variance stabilizing transformation for the density estimation problem, so that a global roughness penalty seems appropriate to be imposed on $v = f^{\frac{1}{2}}$ rather than f; see Tukey (1972) and Good and Gaskins (1971, 1980). We then equivalently consider v as the parameter of the problem, let it vary over an appropriate Hilbert space and express (1.1) in terms of it alone.

Estimators of the density f, distribution function F and hazard rate r are derived from $u_n$ by

$$f_n(t) = u_n(t)^2$$

$$\tilde{F}_n(t) = \int_0^t f_n(t)dt$$

$$\text{and } r_n(t) = f_n(t)/[1 - \tilde{F}_n(t)].$$

The existence, uniqueness and implicit representation of $u_n$ as an exponential spline with knots at the uncensored observations are derived in

Section 2, where we also discuss the numerical evaluation of the estimator through an efficient algorithm of Klonias and Nash (1983 a,b) based on a truncated Newton method described in Nash (1984).

We establish consistency of the proposed estimators under mild moment and smoothness conditions. We rely on asymptotic results for the Kaplan-Meier estimator by Gill (1983) for consistency in probability, and by Földes and Rejtő (1981) for almost sure consistency. The central proposition establishes consistency of $f_n$ in the Hellinger distance, i.e.,

$$\| u_n - f^{\frac{1}{2}} \|_2 \longrightarrow 0 \quad \text{as } n \to \infty$$

almost surely or in probability under suitable conditions, and determines lower bounds on the rate of convergence in each case. Consequently, we obtain consistency of $u_n$ in the supremum-norm and Sobolev norm, consistency of the density estimator $f_n$ in the $L_1$, $L_2$, supremum, and Sobolev norms, and, uniform convergence of the hazard rate estimator $r_n$ on bounded intervals.

The assumptions, statements, and proofs of the consistency and rate of convergence bounds are presented in section 3. Auxiliary lemmas, which provide bounds on integrals needed to establish consistency in the Hellinger distance, are proved in section 4, including a result regarding the entropy of continuous distributions.

## 2. Formulation of the Estimator

The Random Censorship Model: Let $X_1, X_2, \ldots, X_n$ be independent positive random variables with common density function f and cumulative distribution function F. Let $Y_1, Y_2, \ldots, Y_n$ be independent positive random variables, representing censoring times, with common distribution function G which may be discontinuous or defective. The random variables $Y_1, Y_2, \ldots, Y_n$ are assumed to be independent of $X_1, X_2, \ldots, X_n$. The observations are $\{(Z_i, \delta_i): i = 1, 2, \ldots, n\}$, defined by

$$Z_i = X_i \wedge Y_i$$

$$\text{and} \quad \delta_i = I\{X_i \le Y_i\},$$

where $\wedge$ denotes minimum and $I\{A\}$ denotes the indicator random variable of the event A. Denote the distribution function of $\{Z_i\}$ by H, which is given by

$$1 - H = (1-F)(1-G).$$

Define $\tau_F = \sup\{t: F(t) < 1\}$, with $\tau_G$ and $\tau_H$ defined similarly.

The product-limit estimator $\hat{F}_n$ is given by

$$1 - \hat{F}_n(t) = \prod_{\{i: Z_{ni} \le t\}} \left(1 - \frac{1}{n-i+1}\right)^{\delta(i)},$$

where $Z_{n1} \le Z_{n2} \le \cdots \le Z_{nn}$ denote the ordered observations $\{Z_i\}$, and $\delta_{n1}, \delta_{n2}, \ldots, \delta_{nn}$ denote the corresponding indicators $\{\delta_i\}$. The Kaplan-Meier estimator has jumps only at the observations for which $\delta_i = 1$, which are called uncensored observations. There are a random number $N_n$ of uncensored observations. We let $T_{n1} \le T_{n2} \le \cdots \le T_{nN_n}$ denote the ordered uncensored observations, and let $w_{ni}/n$ denote the size of the jump of $\hat{F}_n$ at $T_{ni}$.

The Optimization Problem: In the censored data setting, (1.1) suggests estimating f by the maximizer $f_n$ of

$$n \int_A \log f \, d\hat{F}_n - (\alpha/4) \int_A (f'/f)^2 f$$

$$\text{subject to: } \int_A f = 1, \ f \geq 0,$$

or equivalently, see Lemma (3.1) of DeMontricher et al (1975), by $u_n^2 = f_n$, where $u_n$ denotes the maximizer of the following optimization problem:

$$\text{(2.1)} \qquad \max \{n \int_A \log u^2 \, d\hat{F}_n - \alpha \int_A (u')^2, \ u \in H(A)\}$$

$$\text{subject to: } \int_A u^2 = 1$$

$$\text{and} \quad u(T_{ni}) \geq 0, \ \ 1 \leq i \leq N_n,$$

where $\alpha > 0$, $H(A) = \{u \in L_2(A): \ u' \in L_2(A)\}$ and consider the cases $A + \mathbb{R}_+$, $A = \mathbb{R}$. Incorporating the first constraint into the objective function, we consider maximization of

$$\ell_\lambda(u) = n\int_A \log u^2 d\hat{F}_n - \alpha\int_A (u')^2 - \lambda\int_A u^2$$

where $\lambda$ is the Lagrange multiplier corresponding to the constraint. Then the solutions $u_n^{(1)}$, $u_n^{(2)}$ of (2.1) over $\mathbb{R}$ and $\mathbb{R}_+$ respectively are given implicitly by

$$\text{(2.2)} \qquad u_n^{(j)}(t) = \lambda^{-1} \sum_{i=1}^{N_n} w_{ni} \, u_n^{(j)}(T_{ni})^{-1} k_j(t,T_{ni};h), \ j = 1,2,$$

where $\lambda > 0$ is the Lagrange multiplier associated with the first constraint, $h = (\alpha/\lambda)^{1/2}$ and

$$k_1(x,y;h) = h^{-1} e((x-y)/h),$$

$$k_2(x,y;h) = (2h)^{-1}\{e((x-y)/h) + e((x+y)/h)\},$$

where,

$$e(x) = \exp\{-|x|\}/2, \quad x \in \mathbb{R}.$$

Note that $k_1, k_2$ are the kernels of the reproducing kernel Hilbert spaces (RKHS) $H(A)$, endowed with the inner products $\langle u_1, u_2 \rangle = \int_A u_1 u_2 + h^2 \int_A u_1' u_2'$ for $A = \mathbb{R}, \mathbb{R}_+$ respectively. The parameter h plays the role of the "bandwidth" of a kernel estimator and we will equivalently use h rather $\alpha$ as our smoothing parameter. For the consistency results of Section 3 we will let h depend on n, i.e., $h_n = O(n^{-\xi})$, $\xi > 0$. Then $\lambda$, which also depends on n, behaves asymptotically like n, so that $\alpha_n = O(n^{1-2\xi})$. For a development of the consistency of the MPLE's in the uncensored case, with $\alpha$ rather than h the independent parameter, see Klonias (1982).

In the remainder of the paper, $u_n$ refers specifically to $u_n^{(2)}$, $k=k_2$, $A = \mathbb{R}_+$ and H denotes $H(\mathbb{R}_+)$. However, arguments applying to $u_n^{(1)}$ are nearly identical and, in instances when they differ, are slightly simpler.

Existence and Uniqueness:

PROPOSITION 2.3: Let $H_0(A) = \{u \in H(A): u(T_{ni}) \geq 0,\ i = 1,2,\ldots,N_n\}$. For each $\lambda > 0$, there exists a unique maximizer $u_\lambda$ of (2.1) in $H_0(A)$, which is a spline function given implicitly by

$$u_\lambda(t) = \frac{1}{\lambda} \sum_{i=1}^{N_n} \frac{w_{ni}}{u_\lambda(T_{ni})} K(t, T_{ni}; h) \ \forall t \in A.$$

Then $u_n = u_{\lambda_n}$, where $\lambda_n$ is the value of $\lambda$ for which $\|u_\lambda\|_2^2 = 1$.

Proof. The proof relies on Theorem 7 of Appendix I of Tapia and Thompson (1978). The set $H_0(A) = \{u \in H: u(T_{ni}) \geq 0, 1 \leq i \leq N_n\}$ is closed and convex. The second Gâteaux variation of $\ell_\lambda(u)$, given by

$$\nabla^2 \ell_\lambda(u)(\eta, h) = -2\{\sum_{i=1}^{N_n} w_{ni}\, u(T_{ni})^{-2} \eta(T_{ni}) \xi(T_{ni}) + \lambda \langle \eta, \xi \rangle\},$$

$\eta, \xi \in H$, is uniformly negative definite: $\nabla^2 \ell_\lambda(u)(\eta,\eta) \leq -2\lambda\|\eta\|^2$. To establish the existence and uniqueness of the maximizer of $\ell_\lambda(u)$ over $H_0(A)$ by Tapia and Thompson's result, it suffices to show that $\ell_\lambda$ is continuous. Continuity of $\ell_\lambda$ follows from

$$\Big| \|u\| - \|u^*\| \Big| \leq \|u-u^*\|$$

and

$$|u(T_{ni})-u^*(T_{ni})| \leq | < k(\cdot,T_{ni};h), u-u^* > | \leq \|k(\cdot,T_{ni};h)\| \; \|u-u^*\|.$$

Note that the constraints $u(T_{ni}) \geq 0$ cannot be active at the maximum, so the stationary point of the Lagrangian of the problem satisfies

$$\nabla \ell_\lambda(u)(\eta) = 0 \qquad \forall \eta \in H.$$

Equivalently,

$$< \sum_{i=1}^{N_n} w_{ni} \, u(T_{ni})^{-1} \, k(\cdot,T_{ni};h) - \lambda u, \eta > = 0 \qquad \forall \eta \in H,$$

so

$$\|\lambda^{-1} \sum_{i=1}^{N_n} w_{ni} \, u(T_{ni})^{-1} k(\cdot,T_{ni};\ ) - u\| = 0.$$

provides the form of the maximizer implicitly. $\square$

<u>Numerical Evaluation</u>: Since $u_n$ is determined by its values at the uncensored observations, we will obtain $u_n(T_{ni})$ by first solving for

$$q_i = (\lambda h)^{-\frac{1}{2}} u_n(T_{ni})^{-1}, \quad 1 \leq i \leq N_n.$$

Evaluating the implicit form of $u_n$ at $\{T_{ni}\}$, we obtain the following nonlinear

systems of equations

$$(2.3) \qquad q_i^{-1} = \sum_{j=1}^{N_n} w_{nj}\, k(T_{ni},T_{nj};h), \quad i \le i \le N_n.$$

Note that the solution $\underline{q}_i^{-1} = (q_1,\ldots,q_{N_n})^T$ of (2.3) is the unique solution to the following finite dimensional optimization problem:

$$\min\{\underline{q}^T W \ddagger W\underline{q} - \sum_{i=1}^{N_n} w_{ni} \log q_i^2, \quad \underline{q} \in \mathbb{R}_+^{N_n}\},$$

where $W = \mathrm{diag}\{w_{n1},\ldots,w_{nN_n}\}$ and the (i,j)-th entry of the positive definite matrix $\ddagger$ is given by $k(T_{ni},T_{nj};h)$. Existence and uniqueness of the solution to this problem are discussed in Klonias and Nash (1983), who present an efficient algorithm for the numerical evaluation of $\underline{q}$. The parameter $\lambda$ is then obtained from the equation $\int u_n^2 = 1$, i.e.,

$$\lambda = \sum_{i=1}^{N_n} \sum_{j=1}^{N_n} q_i q_j\, (k*k)(T_{ni},T_{nj};h),$$

where * denotes convolution. The values $u_n(T_{ni})$ are then obtained from $\underline{q}$ and $\lambda$.

In the figures that follow we graph the MPLE $f_n^{(1)} \in H(\mathbb{R})$ or $f_n^{(2)} \in H(\mathbb{R}_+)$ - solid lines - against the underlying density f. The data were generated using the IMSL random number generator GGWIB. The X and Y samples were generated consecutively starting with DSEED = 255866175. The sample sizes are n = 120. In Figure 1, Exponentials $\mathcal{E}(\theta)$, with mean $\theta = 1$, have been censored by $\mathcal{E}(3)$; the number of uncensored observations is $N_n = 86$. In Figure 2, Weibulls $W(\alpha,\theta)$, with shape parameter $\alpha = 3$ and mean $\theta = 1$, have been censored by $W(3,2)$; the number of uncensored observations is $N_n = 105$. Note that in the case of the Exponential density $f^{1/2} \in H(\mathbb{R}_+)$ but $f^{1/2} \notin H(\mathbb{R})$ and as expected $f_n^{(2)}$ performs better than $f_n^{(1)}$. On the other hand in the case of the Weibull, $f^{1/2} \in H(\mathbb{R}) \cap H(\mathbb{R}_+)$ and $f_n^{(1)}$ seems to perform better than $f_n^{(2)}$.

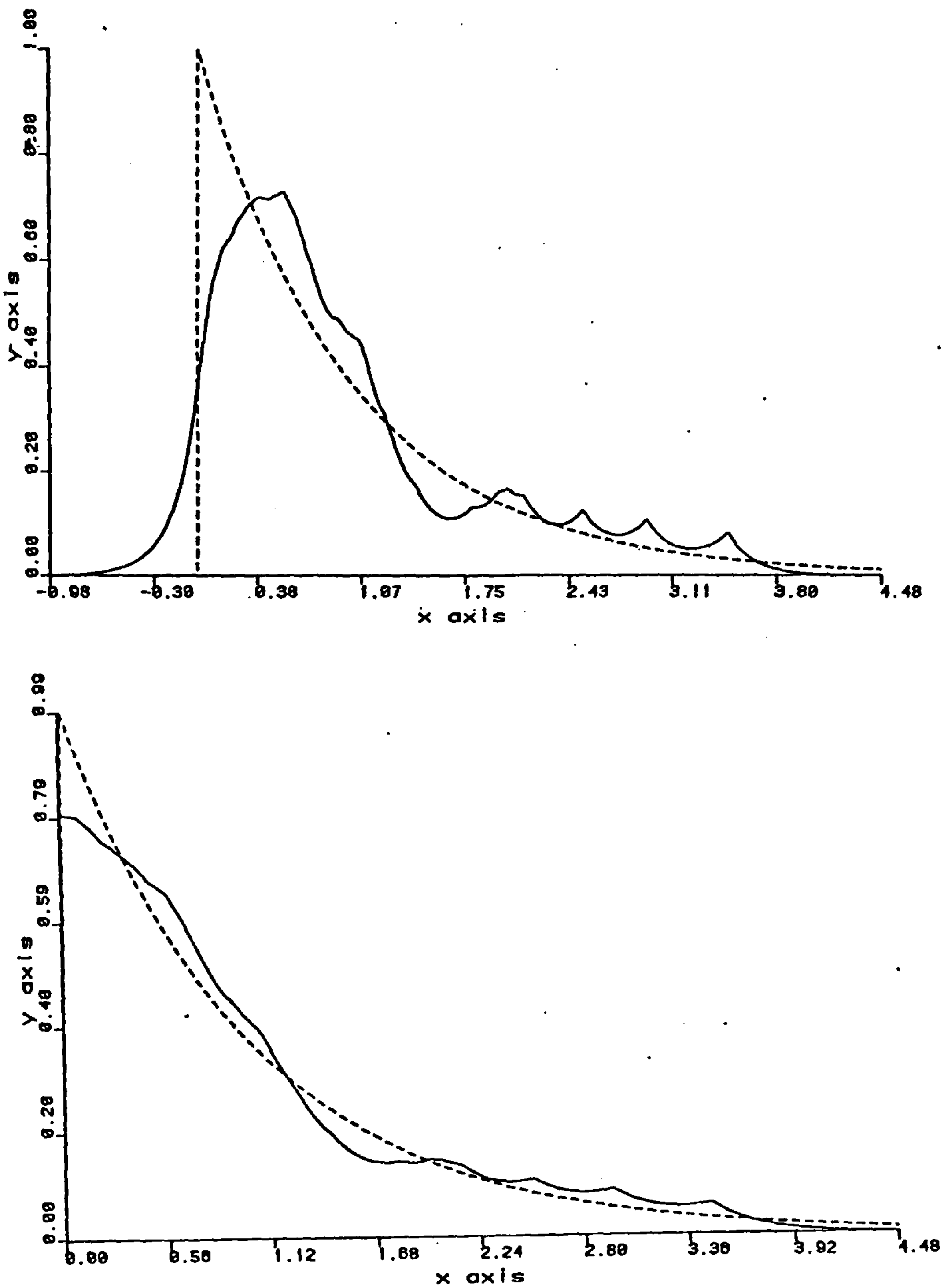



FIG. 1. The solid lines are the estimates $f_n^{(1)} \in H(\mathbb{R})$ (top) with data based $h = .3242$, and $f_n^{(2)} \in H(\mathbb{R}_+)$ with $h = .5$, plotted against the underlying $f$ which is an $E(1)$ density which is censored by an $E(3)$. The data is 120 observations of which 86 are uncensored. Note that $f^{1/2} \in H(\mathbb{R}_+)$ but not in $H(\mathbb{R})$ and $f_n^{(2)}$ performs better than $f_n^{(1)}$.

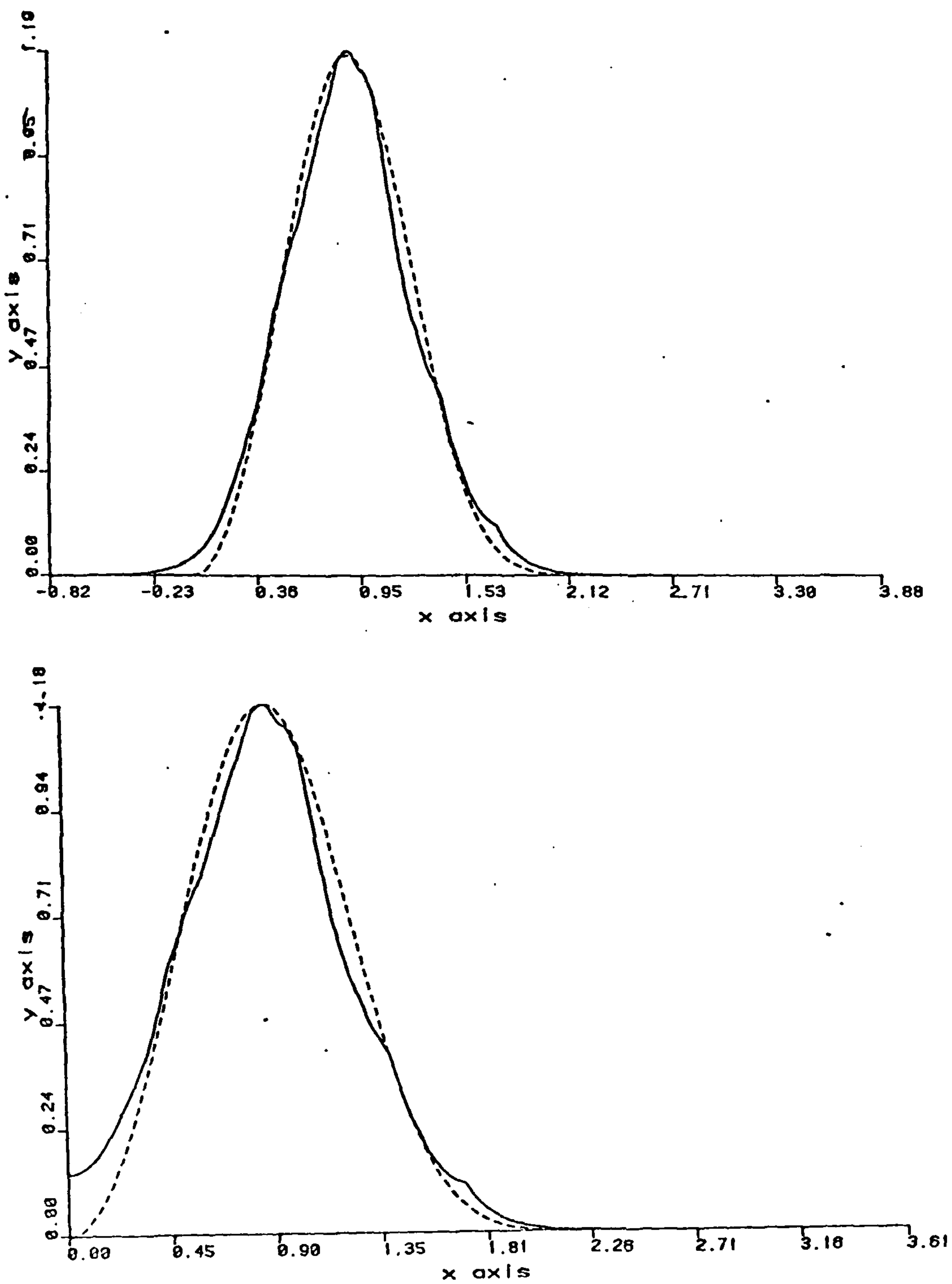



FIG. 2. The solid lines are the estimates $f_n^{(1)} \in H(\mathbb{R})$ (top) with $h = .275$ and $f_n^{(2)} \in H(\mathbb{R}_+)$ with $h = .26$, plotted against the underlying f which is a Weibull(3) with mean 1 density which is censored by a Weibull(3) with mean 2. The data is 120 observations of which 105 are uncensored.
Note that $f^{\frac{1}{2}} \in H(\mathbb{R}) \cap H(\mathbb{R}_+)$ and $f_n^{(1)}$ seems to perform better than $f_n^{(2)}$.

## 3. Consistency

Assumptions: In the proofs of consistency of our estimators, we make use of the following assumptions:

A1: There exists $\tau > 2$ such that $E[X_1^\tau] < \infty$ .

A2: $\int_0^{\tau_H} [1-G(s)]^{-1} \, dF(s) < +\infty$ .

A3: $\|(f^{\frac{1}{2}})'\|_2^2 = \frac{1}{4} \int \left(\frac{f'}{f}\right)^2 dF < +\infty$ .

A4: $f'$ changes sign finitely many times.

A5: $\tau_F < \tau_G \leq +\infty$ .

Assumptions A1 - A4 suffice to establish consistency in probability, employing the results of Gill (1983) on the weak convergence of the Kaplan-Meier estimator, which require A2. To obtain almost sure consistency, we use A5 in place of A1 and A2, in order to apply the law of the iterated logarithm for the Kaplan-Meier estimator due to Foldes and Rejtő (1981). The moment condition A1 is used to obtain upper bounds on the maximum of the uncensored observations. A3 arises in the application of a bound on the entropy, proved in section 4. A4 arises as a technical condition in a law of large numbers for integrals relative to the Kaplan-Meier estimator proved in section 4.

The Root-Density Estimator: We now prove a number of consistency results for the root-density estimator $u_n$, and establish lower bounds on the rates of convergence. The consistency of the density estimator $f_n$ is then derived in the following subsection as a corollary of these convergence properties of $u_n$. The key Theorem 3.3 provides consistency of $u_n$ in the $L_2$-norm, relying on a series of lemmas proved in section 4.

We record the following two facts:

(i) *From the implicit form of* $u_n$, *note that*

$$\|u_n\| = 1 + h_n^2 \|u_n'\|_2^2 = (n/\lambda_n)\,\hat{F}_n(+\infty). \tag{3.1}$$

(ii) Since $u_n$ is the maximizer in the optimization problem,

$$\int_0^\infty \log v^2\, d\hat{F}_n - \frac{\lambda_n}{n} h_n^2 \|v'\| \le \int_0^\infty \log u_n^2\, dF_n - \frac{\lambda_n}{n} h_n^2 \|u_n'\|_2^2. \tag{3.2}$$

__Theorem 3.3__: (a) Under assumptions A1 - A4, and $t > \left(\frac{1}{2} - \frac{1}{\tau}\right)/3$, $\|u_n - v\|_2 = O_p(n^{-d})$ for $d < \left(\frac{1}{2} - \frac{1}{\tau} - t\right)/2$.

(b) Under assumptions A3 - A5, and $t > 1/6$

$$\|u_n - v\|_2 = O(n^{-d}) \text{ a.s. for } d < \left(\frac{1}{2} - t\right)/2.$$

__Proof__. By Lemma 5.3 of Klonias (1982),

$$\|u_n - v\|_2^2 \le \int_0^\infty \log v^2\, dF - \int_0^\infty \log u_n^2\, dF$$

$$= \int_0^\infty \log v^2\, d(F-\hat{F}_n) - \int_0^\infty \log v^2\, d\hat{F}_n - \int_0^\infty \log u_n^2\, d(F-\hat{F}_n) + \int_0^\infty \log u_n^2\, d\hat{F}_n$$

which, by (3.2),

$$\le \int_0^\infty \log u_n^2\, d(F_n-\hat{F}) - \int_0^\infty \log v^2\, d(\hat{F}_n - F) + \left(\lambda_n/n\right) h_n^2 \|v'\|_2^2 - \|u_n'\|_2^2\Big),$$

and since by (3.1), $\lambda_n/n \le 1$,

$$\le \int_0^\infty \log u_n^2\, d(\hat{F}_n - F) - \int_0^\infty \log v^2\, d(\hat{F}_n - F) + h_n^2 \|v'\|_2^2.$$

The conclusions follow from Lemmas 4.2 and 4.7, with the rate of convergence being determined by Lemma 4.2. $\square$

The following lemma establishes the rate at which $\lambda_n$ converges to infinity, for use in the proof of Lemma 3.5 below.

Lemma 3.4: $|(\lambda_n/n) - \hat{F}_n(+\infty)| = O(n^{-d})$

(a) in probability for $d < \left(\frac{1}{2} - \frac{1}{\tau} - t\right)/2$,

(b) almost surely for $d < \left(\frac{1}{2} - t\right)/2$,

under the assumptions of Theorem 3.3 (a) or (b), respectively.

Proof: By (3.1),

$$0 \leq \hat{F}_n(+\infty) - (\lambda_n/n) \leq (\lambda_n/n) h_n^2 \|u_n'\|_2^2$$

which by (3.2),

$$\leq \int_0^\infty \log u_n^2 \, d\hat{F}_n - \int_0^\infty \log v^2 \, d\hat{F}_n + h_n^2 \|v'\|_2^2,$$

so the conclusion follows exactly as in Theorem 3.3.

Next, we establish the consistency of our estimate of the Fisher information. This result is needed to prove uniform consistency of $u_n$ in Theorem 3.6.

Lemma 3.5: $n^d\{\|u_n'\|_2^2 - \|v'\|_2^2\} \to 0$

(a) in probability for $d < \left(\frac{1}{2} - \frac{1}{\tau} - 2t\right)$

(b) almost surely for $d < \left(\frac{1}{2} - 2t\right)$

under the assumptions of Theorem 3.3 (a) or (b), respectively.

Proof: Using (3.2), then (1e6.6) in Rao (1973),

$$- \|v'\|_2^2 \le \|u_n'\|_2^2 - \|v'\|_2^2$$

$$\le (n/\lambda_n)h_n^{-2} \{\int_0^\infty \log u_n^2 \, d\hat{F}_n - \int_0^\infty \log v^2 \, d\hat{F}_n\}$$

$$\le (n/\lambda_n)h_n^{-2} \{\int_0^\infty \log u_n^2 \, d(\hat{F}_n - F) - \int_0^\infty \log v^2 \, d(\hat{F}_n - F)\}.$$

The conclusion follows from Lemma 3.4 and Lemmas 3.12 and 3.17. $\square$

We are now ready to establish the uniform consistency of $u_n$.

Theorem 3.6: $\|u_n - v\|_\infty = O(n^{-d})$

(a) in probability for $d < \left(\frac{1}{2} - \frac{1}{\tau} - 2t\right)/2$, $t > \left(\frac{1}{2} - \frac{1}{\tau}\right)/3$

(b) almost surely for $d < \left(\frac{1}{2} - 2t\right)/2$, $t > 1/6$

under the assumptions of Theorem 3.3(a) or (b), respectively.

Proof: Let $u^*(x) = u(|x|)$ for $x \in \mathbb{R}$. Note that $\|u^*\|_2^2 = 2\|u\|_2^2$, $\|u^{*\prime}\|_2^2 = 2\|u'\|$ and $\|u^*\|_\infty = \|u\|_\infty$. Then, as in Klonias (1982), for each $x \in \mathbb{R}$,

$$|u_n(x) - v(x)|^2 = |u_n^*(x) - v^*(x)|^2$$

$$\le h_n^{-1} \|u_n - v\|_2^2 + h_n\left(\|u_n'\|_2^2 - \|v'\|_2^2\right) + 2h_n\|v'\|_2^2$$

$$+ 2h_n^{\frac{1}{2}}\|v'\|_2 \left\{h_n\left(\|u_n'\|_2^2 - \|v'\|_2^2\right) + h_n\|v'\|_2^2\right\}^{\frac{1}{2}}.$$

The conclusion follows from Theorem 3.3 and Proposition 3.5 (taking $d = -t$).

Theorem 3.7: $\|u_n' - v'\|_2 = O(n^{-d})$

(a) in probability for $d < \left(\frac{1}{2} - \frac{1}{\tau} - 2t\right)/4$

(b) almost surely for $d < \left(\frac{1}{2} - 2t\right)/4$

under the assumptions of Theorem 3.3 (a) or (b), respectively.

Proof: By computation, then integration by parts,

$$\|u_n' - v'\|_2^2 = \|u_n'\|_2^2 - \|v'\|_2^2 + 2\int_0^\infty v'(v'-u_n')$$

$$= \|u_n'\|_2^2 - \|v'\|_2^2 - 2v'(0)[v(0)-u_n(0)] + 2\int_0^\infty v''(u_n-v)$$

which, by $\|g\|_\infty^2 \le \|g\|_2 \|g'\|_2$ (see Klonias (1981)) and the Cauchy-Schwartz inequality

$$\le \|u_n'\|_2^2 - \|v'\|_2^2 + 2\left(\|v'\|_2\|v''\|_2\right)^{\frac{1}{2}} \|u_n'-v\|_\infty + 2\|v''\|_2 \|u_n-v\|_2.$$

**The Density, Distribution, and Hazard Function Estimators:**

Using Lemma 4.1 of Klonias (1982) with Theorems 3.3, 3.6, and 3.7, we obtain consistency of the density estimator $f_n$ in various norms:

Theorem 3.8: Under the assumption of Theorem 3.3 (a) or (b)

(i) $\|f_n-f\|_1$ and $\|f_n-f\|_2$ both converge to zero in probability or almost surely with the rates of Theorem 3.3 (a) or (b) respectively.

(ii) $\|f_n-f\|_\infty$ converges to zero in probability or almost surely with the rates of Theorem 3.6 (a) or (b) respectively.

(iii) $\|f_n-f\|_H$ converges to zero in probability or almost surely with the rates $d < \left(\frac{1}{2} - \frac{1}{\tau} - 4t\right)/4$ or $d < \left(\frac{1}{2} - 4t\right)/4$ respectively.

As corollaries of Theorem 3.8, we establish the uniform consistency of the induced estimates of F and r. Note that

$$\|\tilde{F}_n - F\|_\infty \le \|f_n - f\|_1$$

and hence, by the Theorem 3.8 (i), we conclude:

Corollary 3.9: Under the assumptions of Theorem 3.3 (a) or (b),

$$\|\tilde{F}_n - F\|_\infty = 0(n^{-d})$$

in probability or almost surely with the values of $d > 0$ given in Theorem 3.3 (a) or (b) respectively.

For the proof of the consistency of the induced hazard rate estimator, note that

$$|r_n(t) - r(t)| = \left|[1-\tilde{F}_n(t)]^{-1} f_n(t) - [1-F(t)]^{-1} f(t)\right|$$
$$\leq \{[1-\tilde{F}_n(t)][1-F(t)]\}^{-1}\{\|f_n - f\|_\infty + \|f\|_\infty \|\tilde{F}_n - F\|_\infty\},$$

so by Corollary 3.9 and Theorem 3.8, we have the following result:

Corollary 3.10: Let $I = [0, F^{-1}(1-\varepsilon)]$ for a fixed $\varepsilon > 0$. Then

$$\sup_{t\varepsilon I} |r_n(t) - r(t)| = 0(n^{-d})$$

(a) in probability for $d < \left(\frac{1}{2} - \frac{1}{\tau} - 2t\right)/2$

(b) almost surely for $d < \left(\frac{1}{2} - 2t\right)/2$

under the assumptions of Theorem 3.3 (a) or (b), respectively.

## 4. Auxiliary Lemmas

We now establish bounds for the integrals in the proof of Proposition 3.3. We will use bounds on the rate of convergence of the Kaplan-Meier estimator and on maxima of i.i.d. random variables, which are stated in the following remarks:

R1: If $\tau_F = +\infty$, for any $\beta < \frac{1}{2}$,

$$\|\hat{F}_n - F\|_\infty = O_p(n^{-\beta})$$

as $n \to \infty$, by an application of Theorem 2.1 of Gill (1983).

R2: If F and G are continuous distribution functions such that $\tau_F < \tau_G \leq +\infty$, then

$$\|\hat{F}_n - F\|_\infty = O\left(\left(\frac{\log\log n}{n}\right)^{\frac{1}{2}}\right) \quad \text{a.s.}$$

[See Földes and Rejtő (1981).]

R3: If $E[X_1^\tau] < \infty$, then $X_{nn} = O(n^p)$ almost surely for any $p > 1/\tau$. This fact follows from a tail probability bound, the Borel-Cantelli Lemma, and monotonicity of $X_{nn}$ in n.

Lemma 4.1: Assume that A2 holds and $\tau_F = +\infty$. Then

$$n^d \int_{X_{nn}}^{\infty} [1-F(t)]\,dt \longrightarrow 0 \text{ a.s.}$$

for all $d < 1 - \tau^{-1}$.

Proof: Let $W_n = \int_{X_{nn}}^{\infty} [1-F(t)]\,dt$. Compute

$$\begin{aligned} E[W_n] &= \int_0^\infty \left[\int_s^\infty [1-F(t)]\,dt\right] nF(s)^{n-1} f(s)\,ds \\ &= \int_0^\infty [1-F(t)]F(t)^n\,dt \\ &\leq \int_0^\infty [1-F(t)]e^{-n[1-F(t)]}\,dt. \end{aligned}$$

If $E[X_1^\tau] < \infty$, there exists a constant c such that $1-F(t) \le c/t^\tau$ for all $t > 0$, and thus $F^{-1}(1-c/t^\tau) \le t$ for all $t > 0$. Splitting the interval of integration at $F^{-1}(1-n^{-1}\log n)$, and using these facts to bound the integrand on the resulting intervals, one obtains

$$E[W_n] = O\left(n^{1/\tau-1}\right).$$

Thus, if $d < 1 - 1/\tau$, by Markov's inequality,

$$P[n^d W_n > \varepsilon] = O\left(n^{d+1/\tau-1}\right),$$

yielding convergence in probability. One can obtain almost sure convergence along a geometric subsequence, and extend to the entire sequence by monotonicity.

<u>Lemma</u> <u>4.2</u>: (a) Under the assumptions A1 and A2, as $n \to \infty$

$$h_n n^d \int_0^\infty \log u_n^2 \, d(\hat{F}_n - F) = o_p(1) \text{ for } d < \frac{1}{2} - \frac{1}{\tau}$$

(b) Under assumption A5, as $n \to \infty$

$$h_n (n/\log\log n)^{\frac{1}{2}} \int_0^\infty \log u_n^2 \, d(\hat{F}_n - F) = O(1) \text{ a.s.}$$

<u>Proof</u>: By differentiation in (2.2),

(4.3) $$|u_n'| \le h_n^{-1} u_n \text{ a.s.}$$

By the Cauchy-Schwartz Inequality, $u_n(x) = \langle k(\cdot - x), u_n \rangle$ and $\frac{1}{2} < \frac{\lambda_n}{n} < 1$ as in DeMontricher et al. (1975),

$$\|u_n\|_\infty \le \|k(\cdot - x)\| \; \|u_n\| \le (2k(0)h_n^{-1} \, n/\lambda_n)^{\frac{1}{2}} \le (4k(0)h_n^{-1})^{\frac{1}{2}}.$$

Thus, for $t \ge T_{nN_n}$,

$$u_n(t) \ge \hat{F}_n(+\infty)(2h_n)^{-\frac{1}{2}} e^{-t/h_n},$$

and for n sufficiently large,

(4.4) $$-t/h_n \le \log u_n(t) < 0.$$

Note that $X_{nn} \ge T_{nN_n}$, and $\hat{F}_n(t)$ is constant for $t \ge T_{nN_n}$. Thus, for n sufficiently large,

$$\left|\int_{X_{nn}}^{+\infty} \log u_n^2(t)\, d(\hat{F}_n - F)(t)\right| = \left|\int_{X_{nn}}^{+\infty} \log u_n^2(t)\, dF(t)\right|$$

$$\le 2h_n^{-1} \int_{X_{nn}}^{+\infty} t\, dF(t).$$

By integration by parts,

$$\int_{X_{nn}}^{+\infty} t dF(t) = -\int_{X_{nn}}^{\infty} t d[1-F](t)$$

$$= \int_{X_{nn}}^{\infty} [1-F(t)]dt + X_{nn}[1-F(X_{nn})].$$

Since $\{F(X_i)\}$ are i.i.d. *Uniform* (0,1) random variables,

$$1-F(X_{nn}) = O(n^{-1} \log n) \text{ almost surely.}$$

Combined with the bounds of R1 and Lemma 2.5, this yields

$$\int_{X_{nn}}^{\infty} t\, dF(t) = O(n^{p-1} \log n)$$

for any $p > 1/\tau$. Thus

$$\left|\int_{X_{nn}}^{+\infty} \log u_n^2(t)\, d(\hat{F}_n - F)(t)\right| = O(h_n n^d)$$

for any $d < 1 - 1/\tau$.

To bound the integral from 0 to $X_{nn}$, integrate by parts:

$$\left|\int_0^{X_{nn}} \log u_n^2 \, d(\hat{F}_n - F)\right| = \left|\log u_n^2(X_{nn})[\hat{F}_n(X_{nn}) - F(X_{nn})] - 2\int_0^{X_{nn}} \frac{u_n'}{u_n}(\hat{F}_n - F)\right|$$

$$\leq 4h_n^{-1} X_{nn} \|\hat{F}_n - F\|_\infty$$

$$= O_p(h_n^{-1} n^{p-\beta}) \text{ for } \beta < \tfrac{1}{2} \text{ and } p > 1/\tau,$$

where the inequality follows from (4.3) and (4.4), and the rate from R1 and R3. Thus part (a) is verified for any $d < \frac{1}{2} - 1/\tau$.

For part (b), note that $X_{nn} \to \tau_F < +\infty$ a.s. as $n \to \infty$, and the result follows from R2. □

We now establish the rate of convergence of integrals relative to the Kaplan-Meier estimator.

Proposition 4.5: Let $g: [0,\infty) \to \mathbb{R}$ be a differentiable function. Suppose that $g$ has $M < \infty$ intervals of increase or decrease, and that $\int |g|^\gamma dF < +\infty$ for some $\gamma > 4$. Then, under assumption A2,

$$\left|\int_0^\infty g \, d(\hat{F}_n - F)\right| = O_p(n^{-d}) \text{ for } d < \frac{1}{2} - \frac{2}{\gamma}$$

Under assumption A4, the convergence above is with probability one.

Proof: For each $k \in \mathbb{Z}_+$, let $G_k = \{x \in \mathbb{R}: k - 1 \leq g(x) < k\}$. Letting $m_k$ denote the number of intervals of increase or decrease of $g$ which intersect $G_k$, may be partitioned as $G_k = \bigcup_{i=1}^{m_k} G_{k,i}$, where $g$ is monotone on each $G_{k,i}$. For any increasing sequence of integers $\{A_n\}$,

$$\left|\int_0^\infty g \, d(\hat{F}_n - F)\right| \leq \sum_{|k| \leq A_n} \left|\int_{G_k} g \, d(\hat{F}_n - F)\right| + \left|\int_{\{x: |g(x)| > A_n\}} g \, d(\hat{F}_n - F)\right|.$$

To bound the sum, integrate by parts over each $G_{k,i}$:

$$\left|\int_{G_k} g \; d(F_n - F)\right| = \sum_{i=1}^{m_k} \left| [g(x)(\hat{F}_n - F)(x)]_{G_{k,i}} - \int_{G_{k,i}} g'(x)(\hat{F}_n - F)(x)dx \right|$$

$$\leq \|\hat{F}_n - F\|_\infty \sum_{i=1}^{m_k} (2k + \int_{G_{k,i}} |g'(x)|dx)$$

$$= \|\hat{F}_n - F\|_\infty \sum_{i=1}^{m_k} (2k + \left|\int_{G_{k,i}} g'(x)dx\right|)$$

$$\leq \|\hat{F}_n - F\|_\infty \sum_{i=1}^{m_k} (2k + 1)$$

$$\leq M(2k+1) \|\hat{F}_n - F\|_\infty .$$

Thus, the sum is $O_p(A_n^2 n^{-\beta})$ for any $\beta < \frac{1}{2}$, using R1 to bound $\|\hat{F}_n - F\|_\infty$. Use of R2 provides the corresponding almost sure result. To obtain convergence of the sum to zero, we choose $A_n$ of the form $A_n = [n^\eta]$, and thus require $\eta < \beta/2$.

To bound the remaining integral, first show that

$$\int_{\{x:|g(x)|>A_n\}} g \; d\hat{F}_n = 0$$

for n sufficiently large. For $\eta > 1/\gamma$,

$$P(\int_{\{x:|g(x)|>A_n\}} |g| d\hat{F}_n \neq 0 \text{ i.o.}) = P(\exists i \in \{1,\dots,N_n\}: |g(T_{ni})| \geq A_n \text{ i.o.})$$

$$\leq P(\exists i \in \{1,\dots,n\} : |g(X_{ni})| \geq A_n \text{ i.o.})$$

$$= P(\max_{1\leq i\leq n} |g(X_i)| \geq A_n \text{ i.o.}) = 0$$

by R3, since $E|g(X)|^\gamma < \infty$.

Thus, we need only bound $\int_{\{x:|g(x)| \geq A_n\}} g(x) \; dF(x)$ for $n$ sufficiently large. Let $F_g$ denote the distribution function of $|g(X_i)|$. Note that since $E|g(X_1)|^{\gamma} < \infty$, $y^{\gamma}(1-F_g(y)) \to 0$ as $y \to \infty$, so for $y$ sufficiently large, $1-F_g(y) \leq y^{-\gamma}$. Integrating by parts and applying this bound twice,

$$\int_{\{x:|g(x)| \geq A_n\}} |g| \; dF = \int_{A_n}^{\infty} y \; dF_g(y) = A_n(1-F_g(A_n)) + \int_{A_n}^{\infty} [1-F_g(y) \; dy$$

$$\leq A_n^{-\gamma+1} + \frac{1}{\gamma-1} A_n^{-\gamma+1} = O(n^{\eta(\gamma-1)}).$$

For each $\varepsilon > 0$, we can obtain the rate of convergence $d = \frac{1}{2} - \frac{2}{\gamma} - 3\varepsilon$ by letting $\eta = \frac{1}{8} + \varepsilon$ and $\beta = \frac{1}{2} - \varepsilon$. $\square$

The following result for continuous distributions is an analog of the result of Keilson (1971) that a positive integer-valued random variable with a finite moment of any positive order has finite entropy.

Proposition 4.6: Let $f$ be a probability density function such that $||f||_{\infty} < +\infty$ and $\int_{-\infty}^{\infty} x^{\tau} f(x) dx < +\infty$ for some $\tau > 0$. Then for all $\gamma > 0$,

$$\int f(x) |\log f(x)|^{\gamma} dx < +\infty.$$

Proof: Note that $x|\log x|$ is strictly increasing on $(0, 1/e)$. Thus, given $\varepsilon > 0$, $x^{\varepsilon}|\log x^{\varepsilon}| \leq 1/e$ for $0 < x < e^{-1/\varepsilon}$. Equivalently, for $\gamma > 0$,

$$|\log x|^{\gamma} \leq (e\varepsilon)^{-\gamma} x^{-\varepsilon\gamma} \quad \text{for } 0 < x < e^{-1/\varepsilon}.$$

On $D = \{x \in \mathbb{R}: f(x) < e^{-1/\varepsilon}\}$,

$$f(x)|\log f(x)|^{\gamma} \leq (e\varepsilon)^{-\gamma} f(x)^{1-\varepsilon\gamma},$$

so

$$\int_D f|\log f|^{\gamma} \leq (e\varepsilon)^{-\gamma} \int_D f^{1-\varepsilon\gamma}$$

Let $S = \{x \in D : f(x) \leq x^{-(\tau+1)}\}$. On $S$, $f(x)^{-\varepsilon\gamma} < x^{(\gamma+1)\varepsilon\gamma}$,

so

$$\int_D f^{1-\varepsilon\gamma} \leq \int_S x^{-(\tau+1)(1-\varepsilon\gamma)}dx + \int_{D\setminus S} x^{(\tau+1)\varepsilon\gamma} f(x)dx$$

For $\varepsilon$ sufficiently small, $(\tau+1)(1-\varepsilon\gamma) > 1$ and $(\tau+1)\varepsilon\gamma < \tau$, so the right side is finite. Then

$$\int f|\log f|^{\gamma} \leq (|\log||f||_{\infty}| \vee \tfrac{1}{\varepsilon}) \int_{D^c} f(x)dx + (e\varepsilon)^{-\gamma}\int_D f(x)^{1-\varepsilon\gamma}dx < +\infty. \qquad \square$$

To obtain consistency for the estimator $u_n$, we apply Proposition 4.5 to the function $g = \log f$. By Proposition 4.6, if $E|X_1|^{\delta} < \infty$ for any $\delta > 0$, then $\int f(x)|\log f(x)|^{\gamma}dx < \infty$ for every $\gamma > 0$. Noting that $||f||_{\infty} \leq ||(f^{1/2})'||_2^2$ as in Klonias (1981), we obtain the following corollary.

Lemma 4.7: (a) Under assumptions A1-A4

$$\left|\int_0^{\infty} \log f \, d(\hat{F}_n - F)\right| = o_p(n^{-d}) \quad \text{for } d < 1/2.$$

(b) Under assumptions A3-A5, the convergence in part (a) is almost sure.

Acknowledgment: The authors thank Robert J. Serfling for his advice and helpful comments.

## 5. References


DeMontricher, G.F., Tapia, R.A., and Thompson, J.R. (1975) Nonparametric maximum likelihood estimation of probability densities by penalty function methods. Ann. Statist. 3, 1329-1348.

Földes, A., and Rejtő, L. (1981) A LIL type result for the product limit estimator. Z.F.W. 56, 75-86.

Gill, R. (1983) Large sample behavior of the product-limit estimator on the whole line. Ann. Statist. 11, 49-58.

Good, I.J., and Gaskins, R.A. (1971) Nonparametric roughness penalties for probability densities. Biometrika 58, 255-277.

Good, I.J., and Gaskins, R.A. (1980) Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data. J. Amer. Statist. Assoc. 75, 42-73.

Kaplan, E.L., and Meier, P. (1958) Nonparametric estimation from incomplete observation. J. Amer. Statist. Assoc. 53, 457-481.

Klonias, V.K. (1981) An upper bound to the probability density function in terms of the Fisher information. Johns Hopkins University, Department of Mathematical Sciences, Technical Report No. 336.

Klonias, V.K. (1982) Consistency of two nonparametric maximum penalized likelihood estimators of the probability density function. Ann. Statist. 10, 811-824.

Klonias, V.K. (1984) On a class of nonparametric density and regression estimators. Ann. Statist. 12, 1263-1284.

Klonias, V.K., and Nash, S.G. (1983a) On the computation of a class of maximum penalized likelihood estimators of the probability density function. Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface, North-Holland, 310-314.

Klonias, V.K., and Nash, S.G. (1983b) On the numerical evaluation of a class of nonparametric density and regression estimators. Johns Hopkins University Mathematical Sciences Department. Technical Report No. 396.

Lubecke, A.M., and Padgett, W.J. (1985) Nonparametric maximum penalized likelihood estimation of a density from arbitrarily right-censored observations. Communications in Statistics, to appear.

Nash, Stephen G. (1984) Newton-type minimization via the Lanczos method. SIAM J. Numer. Anal. 21, 770-787.

Rao, C.R. (1973) Linear Statistical Inference and Its Applications, Second edition, John Wiley & Sons, Inc., New York.

Tapia, R.A., and Thompson, J.R. (1978) Nonparametric Probability Density Estimation, Johns Hopkins University Press, Baltimore.

Tukey, John W. (1977) Exploratory Data Analysis, Addison-Wesley.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT CATALOG NUMBER |
|---|---|---|
| ONR No. 85-3 | AD-A163388 | |

| | |
|---|---|
| 4. TITLE<br>On density estimation from censored data by penalized likeluhood methods | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical Report<br>6. PERFORMING ORGANIZATION REPORT NO.<br>Technical Report No. 418 |
| 7. AUTHOR(s)<br>V.K. Klonias, John C. Wierman | 8. CONTRACT OR GRANT NUMBER(s)<br>ONR No. No.00014-79-C-0801 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Department of Mathematical Sciences<br>The Johns Hopkins University<br>Baltimore, Maryland 21218 | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME & ADDRESS<br>Office of Naval Research<br>Statistics and Probability Program<br>Arlington, Virginia 22217 | 12. REPORT DATE<br>June 1985<br>13. NUMBER OF PAGES<br>25 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | 15. SECURITY CLASS (of this report)<br>Unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS
maximum penalized likelihood estimation, density estimation, censored data, Kaplan-Meier estimator, hazard rate

20. ABSTRACT Estimators for the probability density function, cumulative distribution function, and hazard function are proposed in the random censorship setting. The estimators are derived by maximum penalized likelihood methods. Existence, uniqueness, and consistency of the estimates are proved, and computational algorithms are provided.

END

FILMED

2-86

DTIC